SIGGIS Demonstration: Intersection of Social Media Analytics and GeoAI

AMCIS 2020
11 August 2020
Social Media Analytics and GeoAI

Anthony Corso, Ph.D.
Associate Professor of Computing, Software and Data Sciences
California Baptist University
Riverside, CA

Brian Hilton, Ph.D.
Clinical Full Professor
Claremont Graduate University
Director, Advanced GIS Lab
Claremont, CA
Social Media Analytics and GeoAI

• Today, Social Media such as Twitter, Reddit, and Facebook, have become de facto global communication channels to disseminate news, entertainment, and one’s self-revelations.

• This session will demonstrate Social Media preprocessing techniques, the use of Natural Language Processing to augment the data, and geospatial analysis of this data using GeoAI.
Social Media Analytics

• And now, Anthony...
GeoAI

• Brief Discussion
  • Discrete Global Grid Systems
  • Types of Geospatial Data Analytics
  • Types of GeoAI

• Two Examples:
  • “Real-time“, descriptive / diagnostic, spatial-temporal analysis of Tweets
  • Historic, predictive, spatial-temporal analysis of Tweets
Discrete Global Grid Systems

• What is a Discrete Global Grid (DGG)?

• A Discrete Global Grid (DGG) consists of a set of regions that form a partition of the Earth’s surface, where each region has a single point contained in the region associated with it. Each region/point combination is called a cell. Depending on the application, data objects or values may be associated with the regions, points, or cells of a DGG. A Discrete Global Grid System (DGGS) is a series of discrete global grids, usually consisting of increasingly finer resolution grids (though the term DGG is often used interchangeably with the term DGGS).
Discrete Global Grid Systems

Quadrilateral

Hexagonal

Triangular
Discrete Global Grid Systems

• DGGS Resources
  • Southern Terra Cognita Laboratory
  • OGC Specification
  • Uber H3
Discrete Global Grid Systems - H3
Discrete Global Grid Systems - H3

Each hexagon has a unique index value at a specific resolution. At this location, at resolution 8, the hexID = 8829a1d719fffff
At this location, at resolution 9, the hexID = 8929a1d7193ffff
The three points here could be “tagged” with this value
# Discrete Global Grid Systems - H3

<table>
<thead>
<tr>
<th>H3 Resolution</th>
<th>Average Hexagon Area (km²)</th>
<th>Average Hexagon Edge Length (km)</th>
<th>Number of unique indexes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4,250,546.8477000</td>
<td>1,107.712591000</td>
<td>122</td>
</tr>
<tr>
<td>1</td>
<td>607,220.9782429</td>
<td>418.676005500</td>
<td>842</td>
</tr>
<tr>
<td>2</td>
<td>86,745.8540347</td>
<td>158.244655800</td>
<td>5,882</td>
</tr>
<tr>
<td>3</td>
<td>12,392.264821</td>
<td>59.810857940</td>
<td>41,162</td>
</tr>
<tr>
<td>4</td>
<td>1,770.3235517</td>
<td>22.606379400</td>
<td>288,122</td>
</tr>
<tr>
<td>5</td>
<td>252.9033645</td>
<td>8.544408276</td>
<td>2,016,842</td>
</tr>
<tr>
<td>6</td>
<td>36.1290521</td>
<td>3.229482772</td>
<td>14,117,882</td>
</tr>
<tr>
<td>7</td>
<td>5.1612932</td>
<td>1.220629759</td>
<td>98,825,162</td>
</tr>
<tr>
<td>8</td>
<td>0.7373276</td>
<td>0.461354684</td>
<td>691,776,122</td>
</tr>
<tr>
<td>9</td>
<td>0.1053325</td>
<td>0.174375668</td>
<td>4,842,432,842</td>
</tr>
<tr>
<td>10</td>
<td>0.0150475</td>
<td>0.065907807</td>
<td>33,897,029,882</td>
</tr>
<tr>
<td>11</td>
<td>0.0021496</td>
<td>0.024910561</td>
<td>237,279,209,162</td>
</tr>
<tr>
<td>12</td>
<td>0.0003071</td>
<td>0.009415526</td>
<td>1,660,954,464,122</td>
</tr>
<tr>
<td>13</td>
<td>0.0000439</td>
<td>0.003559893</td>
<td>11,626,681,248,842</td>
</tr>
<tr>
<td>14</td>
<td>0.0000063</td>
<td>0.001348575</td>
<td>81,386,768,741,882</td>
</tr>
<tr>
<td>15</td>
<td>0.0000009</td>
<td>0.000509713</td>
<td>569,707,381,193,162</td>
</tr>
</tbody>
</table>
Discrete Global Grid Systems - H3

• What do those resolutions mean?

• For example:
  • Resolution 7: City District
  • Resolution 8: City Neighborhood
  • Resolution 9: 4-8 city blocks
  • Resolution 10: A city block or less
  • Resolution 15: Less than one square meter
Discrete Global Grid Systems - H3 - San Diego

• H3 - San Diego H3 resolution example - Python notebook
• Link
Discrete Global Grid Systems - H3 - San Diego

Example location - Petco Park (San Diego, CA) Google Maps (longitude = -117.157, latitude = 32.708)
Discrete Global Grid Systems - H3 - San Diego

San Diego H3 resolution example
Discrete Global Grid Systems - H3 - San Diego
Discrete Global Grid Systems - H3 - San Diego

San Diego H3 resolution example
Types of Geospatial Data Analytics

4 types of Data Analytics

Value

Prescriptive

Predictive

Diagnostic

Descriptive

What is the data telling you?

Descriptive: What’s happening in my business?
- Comprehensive, accurate and live data
- Effective visualisation

Diagnostic: Why is it happening?
- Ability to drill down to the root-cause
- Ability to isolate all confounding information

Predictive: What’s likely to happen?
- Business strategies have remained fairly consistent over time
- Historical patterns being used to predict specific outcomes using algorithms
- Decisions are automated using algorithms and technology

Prescriptive: What do I need to do?
- Recommended actions and strategies based on champion/challenger testing strategy outcomes
- Applying advanced analytical techniques to make specific recommendations
Types of Geospatial Data Analytics

4 types of Data Analytics

What is the data telling you?

Descriptive: What’s happening in my business?
- Comprehensive, accurate and live data
- Effective visualisation

Diagnostic: Why is it happening?
- Ability to drill down to the root-cause
- Ability to isolate all confounding information

Predictive: What’s likely to happen?
- Business strategies have remained fairly consistent over time
- Historical patterns being used to predict specific outcomes using algorithms
- Decisions are automated using algorithms and technology

Prescriptive: What do I need to do?
- Recommended actions and strategies based on champion/challenger testing strategy outcomes
- Applying advanced analytical techniques to make specific recommendations

Value

Prescriptive
Predictive
Diagnostic
Descriptive

Complexity

www.principa.co.za
Types of Geospatial Data Analytics

What is the data telling you?

Descriptive: *What’s happening in my business?*
- Comprehensive, accurate and live data
- Effective visualisation

Diagnostic: *Why is it happening?*
- Ability to drill down to the root-cause
- Ability to isolate all confounding information

Predictive: *What’s likely to happen?*
- Business strategies have remained fairly consistent over time
- Historical patterns being used to predict specific outcomes using algorithms
- Decisions are automated using algorithms and technology

Prescriptive: *What do I need to do?*
- Recommended actions and strategies based on champion / challenger testing strategy outcomes
- Applying advanced analytical techniques to make specific recommendations
Types of Geospatial Data Analytics

What is the data telling you?

**Descriptive:** What’s happening in my business?
- Comprehensive, accurate and live data
- Effective visualisation

**Diagnostic:** Why is it happening?
- Ability to drill down to the root-cause
- Ability to isolate all confounding information

**Predictive:** What’s likely to happen?
- Business strategies have remained fairly consistent over time
- Historical patterns being used to predict specific outcomes using algorithms
- Decisions are automated using algorithms and technology

**Prescriptive:** What do I need to do?
- Recommended actions and strategies based on champion / challenger testing strategy outcomes
- Applying advanced analytical techniques to make specific recommendations

4 types of Data Analytics

- Value
- Complexity
- Descriptive
- Diagnostic
- Predictive
- Prescriptive
Types of GeoAI

Artificial Intelligence
When a machine is able to mimic human intelligence by having the ability to predict, classify, learn, plan, reason and/or perceive.

Machine Learning
A subset of AI that incorporates math and statistics in order to learn from the data itself, and improve with experience.

Deep Learning
A subset of ML that uses neural networks to solve ever more complex challenges, such as image, audio, and video classification.
Types of GeoAI

- **Artificial Intelligence**: When a machine is able to mimic human intelligence by having the ability to predict, classify, learn, plan, reason and/or perceive.

- **Machine Learning**: A subset of AI that incorporates math and statistics in order to learn from the data itself, and improve with experience.

- **Deep Learning**: A subset of ML that uses neural networks to solve ever more complex challenges, such as image, audio, and video classification.
GeoAI - Machine Learning

Classification

ArcGIS

Clustering

Prediction
GeoAI - Machine Learning

Classification

The process of deciding to which category an object should be assigned based on a training dataset

**Use Case:** Classify impervious surfaces to help effectively prepare for storm and flood events based on the latest high-resolution imagery

*In ArcGIS:* Maximum Likelihood Classification, Random Trees, Support Vector Machine, Forest-based Classification and Regression
GeoAI - Machine Learning

Clustering
The grouping of observations based on similarities of values or locations

**Use Case:** Given the nearly 50,000 reports of traffic between 5pm and 6pm in Los Angeles (from Traffic Alerts by Waze), where are traffic zones that can be used to elicit feedback from current drivers in the area?

**In ArcGIS:** Spatially Constrained Multivariate Clustering, Multivariate Clustering, Density-based Clustering, Image Segmentation, Hot Spot Analysis, Cluster and Outlier Analysis, Space Time Pattern Mining
GeoAI - Machine Learning

Prediction
Using the known to estimate the unknown

**Use Case:** Accurately predict impacts of climate change on local temperature using global climate model data

*In ArcGIS:* Empirical Bayesian Kriging, Areal Interpolation, EBK Regression Prediction, Ordinary Least Squares Regression and Exploratory Regression, Geographically Weighted Regression, Generalized Linear Regression, Forest-based Classification and Regression
Deep Learning: Computer Vision Use Cases

- Image Classification
- Object Detection
- Semantic Segmentation
- Instance Segmentation

GeoAI - Deep Learning
GeoAI - Deep Learning

Object Detection - Swimming Pools

Classification - Land Cover Type
Types of GeoAI

• GeoAI Resources
  
  • [Medium website: GeoAI - thoughts about where AI and GIS intersect](#)
  • [Spatial Analysis and Data Science at the 2020 Esri User Conference](#)
  • [GeoAI: Vertical Use Cases using AI with ArcGIS](#)
  • [Spatial Analysis and Data Science](#)
  • [Geographic Data Science Lab](#)
  • [Geographic Information Systems and Science](#)
  • [Geographic Data Science with PySAL and the PyData Stack](#)
  • [Geocomputation with R](#)
“Real-time”, descriptive / diagnostic, spatial-temporal analysis of Tweets

• Study Area - San Diego, CA

• Spatial Resolution - H3 resolution 7, 8, and 9

• Time Period - late December 2019 (hence, “real-time” in quotes)

• Data Sets
  • Twitter
  • San Diego Calls for Service (public safety data)
“Real-time“, descriptive / diagnostic, spatial-temporal analysis of Tweets

• Workflow (in brief)
  • Tag data (Tweets and Calls for Service) with H3 index values
  • Link Tweets and Calls for Service using H3 index

• Purpose
  • Proof-of-concept linking live data
  • Visualize data using various techniques
  • Examine data in an exploratory / drill-down approach
"Real-time", descriptive / diagnostic, spatial-temporal analysis of Tweets

### Tweets

<table>
<thead>
<tr>
<th>created_at</th>
<th>Polarity</th>
<th>Emotion</th>
<th>Cognition</th>
<th>hex_0</th>
<th>hex_1</th>
<th>hex_2</th>
<th>hex_3</th>
<th>hex_4</th>
<th>hex_5</th>
<th>hex_6</th>
<th>hex_7</th>
<th>hex_8</th>
<th>hex_9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tue Dec 2</td>
<td>Neutral</td>
<td>Neutral</td>
<td>NA</td>
<td>8029f1</td>
<td>81f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
</tr>
<tr>
<td>Tue Dec 2</td>
<td>Positive</td>
<td>Happy</td>
<td>NA</td>
<td>8029f1</td>
<td>81f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
</tr>
<tr>
<td>Tue Dec 2</td>
<td>Positive</td>
<td>Happy</td>
<td>Real</td>
<td>8029f1</td>
<td>81f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
</tr>
<tr>
<td>Tue Dec 2</td>
<td>Positive</td>
<td>Sad</td>
<td>NA</td>
<td>8029f1</td>
<td>81f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
</tr>
<tr>
<td>Tue Dec 2</td>
<td>Negative</td>
<td>Neutral</td>
<td>Fiction</td>
<td>8029f1</td>
<td>81f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
</tr>
<tr>
<td>Tue Dec 2</td>
<td>Negative</td>
<td>Sad</td>
<td>NA</td>
<td>8029f1</td>
<td>81f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
</tr>
<tr>
<td>Tue Dec 2</td>
<td>Neutral</td>
<td>Happy</td>
<td>Fiction</td>
<td>8029f1</td>
<td>81f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
</tr>
<tr>
<td>Tue Dec 2</td>
<td>Positive</td>
<td>Sad</td>
<td>NA</td>
<td>8029f1</td>
<td>81f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
</tr>
<tr>
<td>Wed Dec 2</td>
<td>Neutral</td>
<td>Sad</td>
<td>NA</td>
<td>8029f1</td>
<td>81f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
<td>8829f1</td>
</tr>
</tbody>
</table>

### Calls for Service

<table>
<thead>
<tr>
<th>call_type_description</th>
<th>priority</th>
<th>priority_text</th>
<th>disp0_c0d</th>
<th>description_1</th>
<th>date_time Converted</th>
<th>lat</th>
<th>lon</th>
<th>hex_10</th>
<th>hex_9</th>
<th>hex_8</th>
<th>hex_7</th>
</tr>
</thead>
<tbody>
<tr>
<td>PED STOP/FIELD IN</td>
<td>2</td>
<td>Dispatch as quickly as possible</td>
<td>CANCEL</td>
<td>12/23/2019 00:04</td>
<td>32.74872</td>
<td>-117.106</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
</tr>
<tr>
<td>PED STOP/FIELD IN</td>
<td>2</td>
<td>Dispatch as quickly as possible</td>
<td>CAN</td>
<td>12/23/2019 00:10</td>
<td>32.73822</td>
<td>-117.111</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
</tr>
<tr>
<td>459A BURGLARY ALARM</td>
<td>2</td>
<td>Dispatch as quickly as possible</td>
<td>NO REPORT REQUIRED</td>
<td>12/23/2019 00:12</td>
<td>32.74855</td>
<td>-117.054</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
</tr>
<tr>
<td>SLEEPER SLEEPER</td>
<td>3</td>
<td>Dispatch as quickly as possible</td>
<td>UNFOUNDED</td>
<td>12/23/2019 00:13</td>
<td>32.75441</td>
<td>-117.248</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
</tr>
<tr>
<td>MPSSST TRAFFIC STOP FRO</td>
<td>2</td>
<td>Dispatch as quickly as possible</td>
<td>OTHER</td>
<td>12/23/2019 00:13</td>
<td>32.97004</td>
<td>-117.084</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
</tr>
<tr>
<td>PED STOP/FIELD IN</td>
<td>2</td>
<td>Dispatch as quickly as possible</td>
<td>NO REPORT REQUIRED</td>
<td>12/23/2019 00:15</td>
<td>32.77491</td>
<td>-117.206</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
</tr>
<tr>
<td>NA~C NARCOTICS ACTIV</td>
<td>2</td>
<td>Dispatch as quickly as possible</td>
<td>NO REPORT REQUIRED</td>
<td>12/23/2019 00:15</td>
<td>32.75344</td>
<td>-117.248</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
</tr>
<tr>
<td>SECURITY CHECK</td>
<td>2</td>
<td>Dispatch as quickly as possible</td>
<td>REPORT</td>
<td>12/23/2019 00:17</td>
<td>32.82044</td>
<td>-117.175</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
</tr>
<tr>
<td>459 BURGLARY IN PRO</td>
<td>1</td>
<td>Dispatch immediately - serious</td>
<td>NO REPORT REQUIRED</td>
<td>12/23/2019 00:21</td>
<td>32.73197</td>
<td>-117.154</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
<td>829a411a</td>
</tr>
</tbody>
</table>
“Real-time”, descriptive / diagnostic, spatial-temporal analysis of Tweets
“Real-time”, descriptive / diagnostic, spatial-temporal analysis of Tweets
“Real-time“, descriptive / diagnostic, spatial-temporal analysis of Tweets
Historic, predictive, spatial-temporal analysis of Tweets

• Study Area - San Diego, CA

• Spatial Resolution - H3 resolution 9

• Time Period - late December 2019

• Data Sets
  • Twitter
  • CalEnviroScreen 3.0 (CES3) Indicators in CalEnviroScreen are measures of either environmental conditions, in the case of pollution burden indicators, or health and vulnerability factors for population characteristics indicators.
Historic, predictive, spatial-temporal analysis of Tweets

• **Workflow (in brief)**
  - Tag data (Tweets) with H3 index values
  - H3 - San Diego H3 hexagons example - Python notebook
  - ArcGIS - San Diego tabulate intersect example - Python notebook
  - Append Tweets data set with CES3 Indicators using H3 index

• **Purpose**
  - Examine relationships between “NLP-ed” Tweets and CES3 data
  - Predict Emotion (Happy, Neutral, Sad) based on CES3 Population Characteristics
Historic, predictive, spatial-temporal analysis of Tweets

CES3 Indicators:
- Pollution
  - Exposures
  - Environmental Effects
- Pollution Characteristics
  - Sensitive Populations
  - Socioeconomic Factors
Historic, predictive, spatial-temporal analysis of Tweets

• H3 - San Diego H3 hexagons example - Python notebook
• Link
Historic, predictive, spatial-temporal analysis of Tweets

San Diego H3 hexagons (resolution 7)
Historic, predictive, spatial-temporal analysis of Tweets

San Diego H3 hexagons (resolution 8)
Historic, predictive, spatial-temporal analysis of Tweets

San Diego H3 hexagons (resolution 9)
Historic, predictive, spatial-temporal analysis of Tweets

San Diego H3 hexagons (resolution 10)
Historic, predictive, spatial-temporal analysis of Tweets

CES3 - Poverty Attribute - Census Tracts
Historic, predictive, spatial-temporal analysis of Tweets
Historic, predictive, spatial-temporal analysis of Tweets

CES3 - Poverty Attribute - San Diego H3 hexagons tabulate intersect
Historic, predictive, spatial-temporal analysis of Tweets

• ArcGIS - San Diego tabulate intersect example - Python notebook
• Link
Historic, predictive, spatial-temporal analysis of Tweets

In [4]:
```
import pandas as pd

def read_csv(filename):
    return pd.read_csv(filename)

df = read_csv('gis_analysis\h3_san_diego_7_areas.csv')
df
```

Out[4]:
```
<table>
<thead>
<tr>
<th>OBJECTID</th>
<th>hex_id</th>
<th>ozone</th>
<th>pm</th>
<th>diesel</th>
<th>drink</th>
<th>pest</th>
<th>RSEIhaz</th>
<th>traffic</th>
<th>cleanups</th>
<th>gwthreats</th>
<th>haz</th>
<th>iwb</th>
<th>swis</th>
<th>asthma</th>
<th>cvd</th>
<th>lbw</th>
<th>edu</th>
<th>housingB</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0.046</td>
<td>8.28</td>
<td>2.71</td>
<td>554.17</td>
<td>13.408</td>
<td>108.40</td>
<td>2394.83</td>
<td>24.75</td>
<td>199.75</td>
<td>8.10</td>
<td>11</td>
<td>9.5</td>
<td>22.72</td>
<td>4.48</td>
<td>3.97</td>
<td>1.2</td>
<td>36.2</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0.046</td>
<td>8.28</td>
<td>2.71</td>
<td>554.17</td>
<td>13.408</td>
<td>108.40</td>
<td>2394.83</td>
<td>24.75</td>
<td>199.75</td>
<td>8.10</td>
<td>11</td>
<td>9.5</td>
<td>22.72</td>
<td>4.48</td>
<td>3.97</td>
<td>1.2</td>
<td>36.2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0.046</td>
<td>8.28</td>
<td>2.71</td>
<td>554.17</td>
<td>13.408</td>
<td>108.40</td>
<td>2394.83</td>
<td>24.75</td>
<td>199.75</td>
<td>8.10</td>
<td>11</td>
<td>9.5</td>
<td>22.72</td>
<td>4.48</td>
<td>3.97</td>
<td>1.2</td>
<td>36.2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0.453</td>
<td>7.44</td>
<td>2.42</td>
<td>907.13</td>
<td>3.031</td>
<td>129.83</td>
<td>682.96</td>
<td>12.00</td>
<td>3.00</td>
<td>0.15</td>
<td>7</td>
<td>0.2</td>
<td>30.39</td>
<td>6.63</td>
<td>4.14</td>
<td>10.6</td>
<td>11.1</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>0.046</td>
<td>8.28</td>
<td>2.71</td>
<td>554.17</td>
<td>13.408</td>
<td>108.40</td>
<td>2394.83</td>
<td>24.75</td>
<td>199.75</td>
<td>8.10</td>
<td>11</td>
<td>9.5</td>
<td>22.72</td>
<td>4.48</td>
<td>3.97</td>
<td>1.2</td>
<td>36.2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>3037</td>
<td>3038</td>
<td>0.048</td>
<td>8.70</td>
<td>0.53</td>
<td>624.70</td>
<td>0.274</td>
<td>11.45</td>
<td>199.96</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>9</td>
<td>0.0</td>
<td>21.81</td>
<td>5.82</td>
<td>4.96</td>
<td>14.3</td>
<td>14.0</td>
</tr>
<tr>
<td>3038</td>
<td>3039</td>
<td>0.055</td>
<td>7.38</td>
<td>0.17</td>
<td>1008.75</td>
<td>0.419</td>
<td>6.67</td>
<td>89.93</td>
<td>0.00</td>
<td>8.00</td>
<td>0.00</td>
<td>11</td>
<td>12.4</td>
<td>21.73</td>
<td>4.71</td>
<td>3.16</td>
<td>12.2</td>
<td>11.2</td>
</tr>
<tr>
<td>3039</td>
<td>3040</td>
<td>0.055</td>
<td>7.38</td>
<td>0.17</td>
<td>1008.75</td>
<td>0.419</td>
<td>6.67</td>
<td>89.93</td>
<td>0.00</td>
<td>8.00</td>
<td>0.00</td>
<td>11</td>
<td>12.4</td>
<td>21.73</td>
<td>4.71</td>
<td>3.16</td>
<td>12.2</td>
<td>11.2</td>
</tr>
<tr>
<td>3040</td>
<td>3041</td>
<td>0.055</td>
<td>7.38</td>
<td>0.17</td>
<td>1008.75</td>
<td>0.419</td>
<td>6.67</td>
<td>89.93</td>
<td>0.00</td>
<td>8.00</td>
<td>0.00</td>
<td>11</td>
<td>12.4</td>
<td>21.73</td>
<td>4.71</td>
<td>3.16</td>
<td>12.2</td>
<td>11.2</td>
</tr>
<tr>
<td>3041</td>
<td>3042</td>
<td>0.055</td>
<td>7.38</td>
<td>0.17</td>
<td>1008.75</td>
<td>0.419</td>
<td>6.67</td>
<td>89.93</td>
<td>0.00</td>
<td>8.00</td>
<td>0.00</td>
<td>11</td>
<td>12.4</td>
<td>21.73</td>
<td>4.71</td>
<td>3.16</td>
<td>12.2</td>
<td>11.2</td>
</tr>
</tbody>
</table>

3042 rows × 24 columns
Historic, predictive, spatial-temporal analysis of Tweets

```python
In [4]:
pd.set_option('max_columns', None)
# pd.set_option("max_rows", None)
df = pd.read_csv('gis_analysis\h3_san_diego_7_areas.csv')
df
```

```
<table>
<thead>
<tr>
<th>asel</th>
<th>drink</th>
<th>pest</th>
<th>RSEL Haz</th>
<th>traffic</th>
<th>cleanups</th>
<th>gwhaz</th>
<th>hz</th>
<th>iwb</th>
<th>swis</th>
<th>asthma</th>
<th>cvd</th>
<th>lbw</th>
<th>edu</th>
<th>housingB</th>
<th>ling</th>
<th>pov</th>
<th>unemp</th>
<th>AREA</th>
<th>PERCENTAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.71</td>
<td>554.17</td>
<td>13.408</td>
<td>108.40</td>
<td>2394.83</td>
<td>24.75</td>
<td>199.75</td>
<td>8.10</td>
<td>10.5</td>
<td>22.72</td>
<td>4.48</td>
<td>3.97</td>
<td>1.2</td>
<td>36.2</td>
<td>2.2</td>
<td>40.4</td>
<td>15.5</td>
<td>8.383652e+06</td>
<td>99.999009</td>
<td></td>
</tr>
<tr>
<td>2.71</td>
<td>554.17</td>
<td>13.408</td>
<td>108.40</td>
<td>2394.83</td>
<td>24.75</td>
<td>199.75</td>
<td>8.10</td>
<td>10.5</td>
<td>22.72</td>
<td>4.48</td>
<td>3.97</td>
<td>1.2</td>
<td>36.2</td>
<td>2.2</td>
<td>40.4</td>
<td>15.5</td>
<td>8.380721e+06</td>
<td>100.000000</td>
<td></td>
</tr>
<tr>
<td>2.71</td>
<td>554.17</td>
<td>13.408</td>
<td>108.40</td>
<td>2394.83</td>
<td>24.75</td>
<td>199.75</td>
<td>8.10</td>
<td>10.5</td>
<td>22.72</td>
<td>4.48</td>
<td>3.97</td>
<td>1.2</td>
<td>36.2</td>
<td>2.2</td>
<td>40.4</td>
<td>15.5</td>
<td>7.104579e+06</td>
<td>84.715167</td>
<td></td>
</tr>
<tr>
<td>2.42</td>
<td>907.13</td>
<td>3.031</td>
<td>129.83</td>
<td>682.96</td>
<td>12.00</td>
<td>3.00</td>
<td>0.15</td>
<td>7</td>
<td>0.2</td>
<td>30.39</td>
<td>6.63</td>
<td>4.14</td>
<td>10.6</td>
<td>11.1</td>
<td>3.1</td>
<td>31.3</td>
<td>4.5</td>
<td>1.231852e+06</td>
<td>15.284833</td>
</tr>
<tr>
<td>2.71</td>
<td>554.17</td>
<td>13.408</td>
<td>108.40</td>
<td>2394.83</td>
<td>24.75</td>
<td>199.75</td>
<td>8.10</td>
<td>10.5</td>
<td>22.72</td>
<td>4.48</td>
<td>3.97</td>
<td>1.2</td>
<td>36.2</td>
<td>2.2</td>
<td>40.4</td>
<td>15.5</td>
<td>8.319281e+06</td>
<td>99.198364</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0.63</td>
<td>624.70</td>
<td>6.274</td>
<td>11.45</td>
<td>199.96</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>9</td>
<td>0.0</td>
<td>21.81</td>
<td>5.82</td>
<td>4.96</td>
<td>14.3</td>
<td>14.0</td>
<td>3.0</td>
<td>38.9</td>
<td>9.0</td>
<td>8.407951e+04</td>
<td>1.002178</td>
</tr>
<tr>
<td>0.17</td>
<td>1008.75</td>
<td>0.419</td>
<td>6.67</td>
<td>89.93</td>
<td>0.00</td>
<td>8.00</td>
<td>0.00</td>
<td>11</td>
<td>12.4</td>
<td>21.73</td>
<td>4.71</td>
<td>3.16</td>
<td>12.2</td>
<td>11.2</td>
<td>2.4</td>
<td>47.0</td>
<td>18.5</td>
<td>8.305597e+06</td>
<td>98.997823</td>
</tr>
<tr>
<td>0.17</td>
<td>1008.75</td>
<td>0.419</td>
<td>6.67</td>
<td>89.93</td>
<td>0.00</td>
<td>8.00</td>
<td>0.00</td>
<td>11</td>
<td>12.4</td>
<td>21.73</td>
<td>4.71</td>
<td>3.16</td>
<td>12.2</td>
<td>11.2</td>
<td>2.4</td>
<td>47.0</td>
<td>18.5</td>
<td>8.392271e+06</td>
<td>100.000000</td>
</tr>
<tr>
<td>0.17</td>
<td>1008.75</td>
<td>0.419</td>
<td>6.67</td>
<td>89.93</td>
<td>0.00</td>
<td>8.00</td>
<td>0.00</td>
<td>11</td>
<td>12.4</td>
<td>21.73</td>
<td>4.71</td>
<td>3.16</td>
<td>12.2</td>
<td>11.2</td>
<td>2.4</td>
<td>47.0</td>
<td>18.5</td>
<td>8.394663e+06</td>
<td>100.000003</td>
</tr>
<tr>
<td>0.17</td>
<td>1008.75</td>
<td>0.419</td>
<td>6.67</td>
<td>89.93</td>
<td>0.00</td>
<td>8.00</td>
<td>0.00</td>
<td>11</td>
<td>12.4</td>
<td>21.73</td>
<td>4.71</td>
<td>3.16</td>
<td>12.2</td>
<td>11.2</td>
<td>2.4</td>
<td>47.0</td>
<td>18.5</td>
<td>8.391717e+06</td>
<td>99.999999</td>
</tr>
</tbody>
</table>
```
Historic, predictive, spatial-temporal analysis of Tweets

```
Out[6]:
```
```
<table>
<thead>
<tr>
<th>hex_id</th>
<th>Ozone</th>
<th>PM_2_5</th>
<th>Diesel_PM</th>
<th>Drinking_Water</th>
<th>Pesticide_Use</th>
<th>Toxic_Releases</th>
<th>Traffic</th>
<th>Cleanup_Sites</th>
<th>Groundwater_Threats</th>
<th>Hazardous</th>
<th>Hazardous</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.048000</td>
<td>8.280000</td>
<td>2.710000</td>
<td>554.168097</td>
<td>13.408000</td>
<td>108.390909</td>
<td>2394.829088</td>
<td>24.750000</td>
<td>199.749099</td>
<td>8.1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.048000</td>
<td>8.280000</td>
<td>2.710000</td>
<td>554.170034</td>
<td>13.408000</td>
<td>108.390909</td>
<td>2394.830005</td>
<td>24.750000</td>
<td>199.750000</td>
<td>8.1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.048000</td>
<td>8.151607</td>
<td>2.665674</td>
<td>606.119343</td>
<td>11.821893</td>
<td>111.67539</td>
<td>2133.173519</td>
<td>22.601184</td>
<td>169.677090</td>
<td>6.8</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.048056</td>
<td>9.273286</td>
<td>2.707675</td>
<td>558.999438</td>
<td>13.324914</td>
<td>109.571702</td>
<td>2381.107022</td>
<td>24.847791</td>
<td>198.172781</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.051698</td>
<td>7.596285</td>
<td>2.473956</td>
<td>841.406554</td>
<td>4.961677</td>
<td>125.842873</td>
<td>1001.459408</td>
<td>14.372182</td>
<td>39.606027</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>963</td>
<td>0.055000</td>
<td>7.390000</td>
<td>0.170000</td>
<td>1008.750011</td>
<td>0.419000</td>
<td>6.670000</td>
<td>89.930001</td>
<td>0.000000</td>
<td>8.000000</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>964</td>
<td>0.054930</td>
<td>7.390329</td>
<td>0.173008</td>
<td>1004.901151</td>
<td>0.477678</td>
<td>8.717004</td>
<td>91.032698</td>
<td>0.000000</td>
<td>7.819828</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>965</td>
<td>0.055000</td>
<td>7.380000</td>
<td>0.170000</td>
<td>1008.749997</td>
<td>0.419000</td>
<td>6.670000</td>
<td>89.930000</td>
<td>0.000000</td>
<td>8.000000</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>966</td>
<td>0.055000</td>
<td>7.380000</td>
<td>0.170000</td>
<td>1008.750028</td>
<td>0.419000</td>
<td>6.670000</td>
<td>89.930028</td>
<td>0.000000</td>
<td>8.000000</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>967</td>
<td>0.055000</td>
<td>7.380000</td>
<td>0.170000</td>
<td>1008.749991</td>
<td>0.419000</td>
<td>6.670000</td>
<td>89.929999</td>
<td>0.000000</td>
<td>8.000000</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>968 rows x 21 columns</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

Historic, predictive, spatial-temporal analysis of Tweets

CES3 - Poverty Attribute - San Diego H3 hexagons (resolution 7) tabulate intersect
Historic, predictive, spatial-temporal analysis of Tweets

CES3 - Poverty Attribute - San Diego H3 hexagons (resolution 8) tabulate intersect
Historic, predictive, spatial-temporal analysis of Tweets

CES3 - Poverty Attribute - San Diego H3 hexagons (resolution 9) tabulate intersect
Historic, predictive, spatial-temporal analysis of Tweets

CES3 - Poverty Attribute - San Diego H3 hexagons (resolution 10) tabulate intersect
Historic, predictive, spatial-temporal analysis of Tweets
Historic, predictive, spatial-temporal analysis of Tweets

Resolution 9: 4-8 city blocks

CES3 - Poverty Attribute - San Diego H3 hexagons (resolution 9)
Historic, predictive, spatial-temporal analysis of Tweets

• Forest-based Classification and Regression
Historic, predictive, spatial-temporal analysis of Tweets

• Forest-based Classification and Regression

• Many decision trees are created, called an ensemble or a forest, that are used for prediction.
• Each tree generates its own prediction and is used as part of a voting scheme to make final predictions.
• Final predictions are not based on any single tree but rather on the entire forest.
Historic, predictive, spatial-temporal analysis of Tweets

• Forest-based Classification and Regression

• The use of the entire forest helps avoid overfitting the model to the training dataset,

• as does the use of both a random subset of the training data and a random subset of explanatory variables in each tree that constitutes the forest.
Historic, predictive, spatial-temporal analysis of Tweets

Dependent Variable
Breed
Size
Color
Fur
Ears
Tail
Age
Weight

Independent Variables

explanatory variables
Historic, predictive, spatial-temporal analysis of Tweets

Decision Tree

- Independent Variable
- Dependent Variable

Size
- Color
- Ears
Historic, predictive, spatial-temporal analysis of Tweets

Random subset of data and variables used in each tree
Historic, predictive, spatial-temporal analysis of Tweets
Historic, predictive, spatial-temporal analysis of Tweets

• Forest-based Classification and Regression

Classification

Predict categorical variable

- Presence of disease
- Crime type
- Causes of forest fires
- Species distribution
- Dog breed
Historic, predictive, spatial-temporal analysis of Tweets

- Forest-based Classification and Regression

**Regression**

Predict continuous variable

- Healthcare spending
- Crime rate
- Mortality rate
- Rate of disease
- Sales profits
Historic, predictive, spatial-temporal analysis of Tweets

• Forest-based Classification and Regression - Model Parameters

  • Predict Emotion (Happy, Neutral, Sad) based on CES3 Population Characteristics
  • 90 training / 10 validation split, 100 trees, 100 iterations

<table>
<thead>
<tr>
<th>Model 1 Variables</th>
<th>Model 2 Variables</th>
<th>Model 3 Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployment</td>
<td>Unemployment</td>
<td>Poverty</td>
</tr>
<tr>
<td>Poverty</td>
<td>Poverty</td>
<td>Asthma</td>
</tr>
<tr>
<td>Linguistic Isolation</td>
<td>Housing Burden</td>
<td>Cardiovascular Disease</td>
</tr>
<tr>
<td>Housing Burden</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educational Attainment</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Historic, predictive, spatial-temporal analysis of Tweets

- Forest-based Classification and Regression - Results

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Actual</th>
<th>Predicted</th>
<th>%_Correct</th>
<th>Actual</th>
<th>Predicted</th>
<th>%_Correct</th>
<th>Actual</th>
<th>Predicted</th>
<th>%_Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td>266</td>
<td>176</td>
<td>66.16541353</td>
<td>269</td>
<td>204</td>
<td>75.83643123</td>
<td>269</td>
<td>195</td>
<td>72.49070632</td>
</tr>
<tr>
<td>Neutral</td>
<td>280</td>
<td>421</td>
<td>150.3571429</td>
<td>280</td>
<td>433</td>
<td>154.6428571</td>
<td>280</td>
<td>400</td>
<td>142.8571429</td>
</tr>
<tr>
<td>Sad</td>
<td>261</td>
<td>210</td>
<td>80.45977011</td>
<td>263</td>
<td>175</td>
<td>66.53992395</td>
<td>263</td>
<td>217</td>
<td>82.5095057</td>
</tr>
</tbody>
</table>

Under

Over

Under
Historic, predictive, spatial-temporal analysis of Tweets

- Forest-based Classification and Regression - Results

<table>
<thead>
<tr>
<th>Model 1 Variables</th>
<th>Importance</th>
<th>Model 2 Variables</th>
<th>Importance</th>
<th>Model 3 Variables</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployment</td>
<td>21%</td>
<td>Unemployment</td>
<td>36%</td>
<td>Poverty</td>
<td>34%</td>
</tr>
<tr>
<td>Poverty</td>
<td>18%</td>
<td>Poverty</td>
<td>29%</td>
<td>Asthma</td>
<td>35%</td>
</tr>
<tr>
<td>Linguistic Isolation</td>
<td>17%</td>
<td>Housing Burden</td>
<td>35%</td>
<td>Cardiovascular Disease</td>
<td>31%</td>
</tr>
<tr>
<td>Housing Burden</td>
<td>22%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educational Attainment</td>
<td>21%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Historic, predictive, spatial-temporal analysis of Tweets

• Forest-based Classification and Regression - Model 1 Actual
Historic, predictive, spatial-temporal analysis of Tweets

- Forest-based Classification and Regression - Model 1 Predicted
Live Demo

• Demonstration using ArcGIS Insights

• Demonstration using ArcGIS Pro